Applied A.I. Solutions

Foundations of Data Management

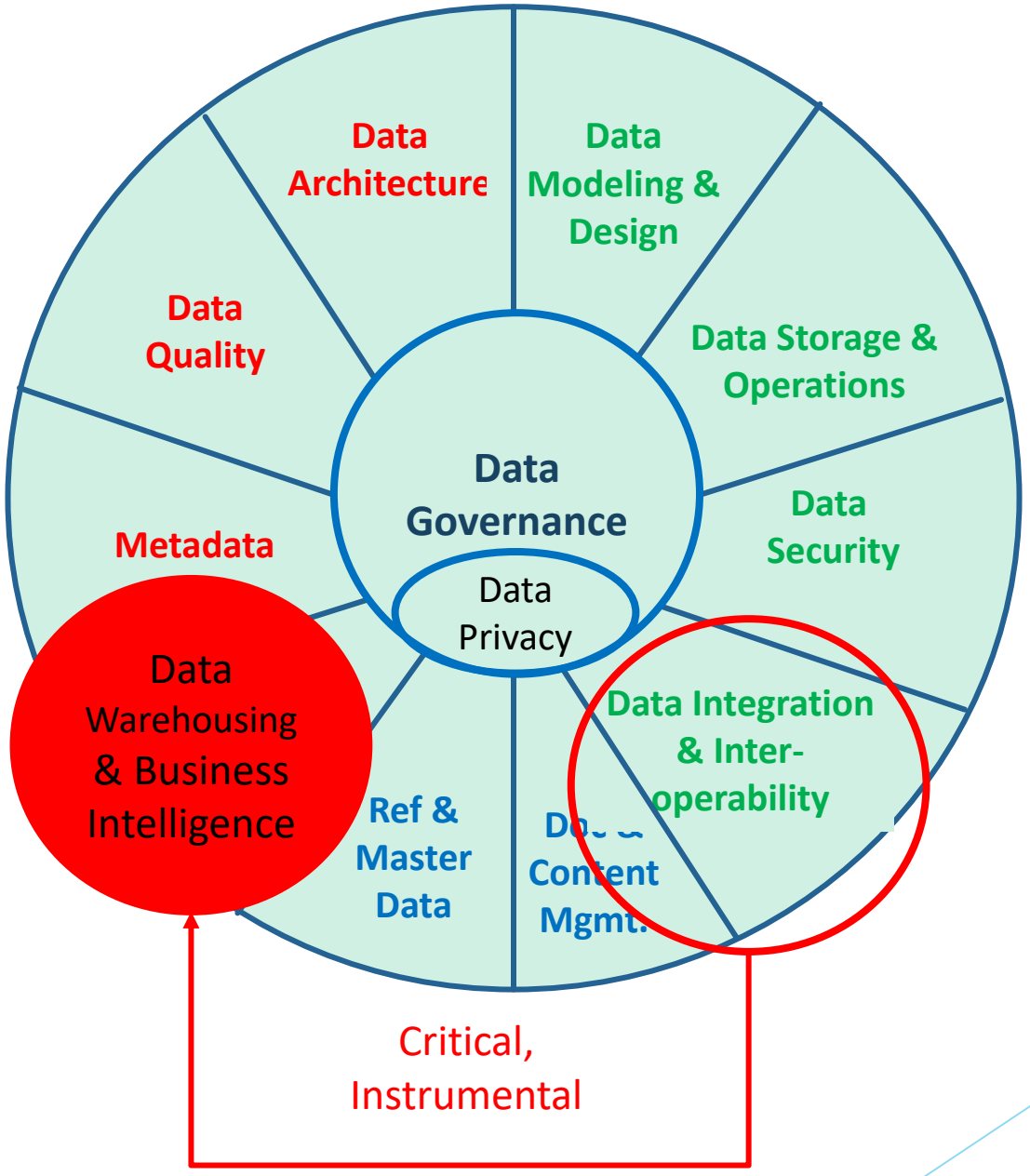
Professor

Daniel Vitaver, EMBA

daniel.vitaver@georgebrown.ca

DATA WAREHOUSING
AND
BUSINESS INTELLIGENCE

The DAMA Wheel



[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

1. INTRODUCTION

Business Intelligence: challenges

- You cannot see the big picture
- You cannot get the right information fast enough for decision-making
- You lack a reliable source of data
- Your personnel cannot collaborate efficiently
- Reporting is hard to pull-out from your systems

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

1. INTRODUCTION

- Technology that enables to **integrate** data from a variety of sources into a common model.

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Business Intelligence Framework

- Data Discovery
- Learning & Knowledge Management
- Informed Decision-making
- Performance Management & Improvement
- Research & Development
- Collaboration & Information Sharing

DW & BI Management Framework

Definition

Planning, implementation, and control processes to provide decision support data for reporting, query, and analysis

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Goals

1. To build and maintain a technical environment and processes needed to deliver integrated data in support to business operations, compliance requirements, and business intelligence activities

Business  **Drivers**

- **Critical Thinking**
  - Decision-making
  - Research and Development (R&D)
  - Performance Improvement
  - Monitoring and Control

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Guiding Principles:

1. Focus on busines goals and priorities
2. Think corporate (architecture), build locally (incrementally)
3. Promote transparency, self-service and collaboration
4. Build Metadata with the warehouse

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Inputs

- Business requirements
- Scalability, operational infrastructure, support requirements
- Data Quality, Security requirements
- IT Strategy, policies and standards
- Master Data and Reference Data
- External data

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Primary Deliverables

- DW and BI architecture
  - Data products
  - Population process
  - Governance activities
  - Lineage dictionary

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Suppliers	Participants	Consumers
• Business executives • Governance body • Enterprise architecture • Data producers • Information consumers • SME	• Sponsor and product owner • Architects and analysts • DW/BI specialists • Project management professionals • Change management professionals	• Information consumers • Customers • Managers and executives

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Business Drivers

- Support operational functions, compliance requirements, BI activities
- Promote evidence-based decision-making process
- Provide and maintain historical data
- BI facilitates insight about the organization, customers, products, etc.
- Improve efficiency and competitive advantage

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Techniques

- Prototypes to drive requirements
- Self service BI
- Audit data

Tools

- Metadata repositories
- Data integration tools
- BI and Analytic applications

Metrics

- Usage metric
- Consumer satisfaction
- Subject area coverage
- Performance

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Essential Concepts

1. Business Intelligence
2. Data Warehouse
3. Data Warehousing
4. Approaches to Data Warehousing

1. Corporate Information factory (Inmon)
2. Dimensional DW (Kimball)
3. DW Architecture Components
4. Type of load processing

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

1. Business Intelligence

   a. Type of **data analysis** aimed at understanding organizational activities and opportunities

   b. Set of **technologies** that support data analysis, and advanced analytics through the **discovery and transformation** of data into meaningful information

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

2. Data Warehouse
   - Integrated decision support database and related software programs used to administer data from a variety of sources

3. Data Warehousing
   - Describes the operational extract, cleansing, transformation, control, and load processes that maintain the data in the DW
   - Enforces business rules, maintains business data relationships

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

4. Approaches to Data Warehousing

   a. **Bill Inmon**: DW is a "subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision-making process" (normalized relational model)

   b. **Ralph Kimball**: DW is "a copy of transaction data specially structured for query and analysis" (dimensional model)

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

5. Corporate Information Factory (Inmon)

- Subject-oriented
- Integrated
- Time variant
- Non-volatile
- Aggregate and detail data
- Historical

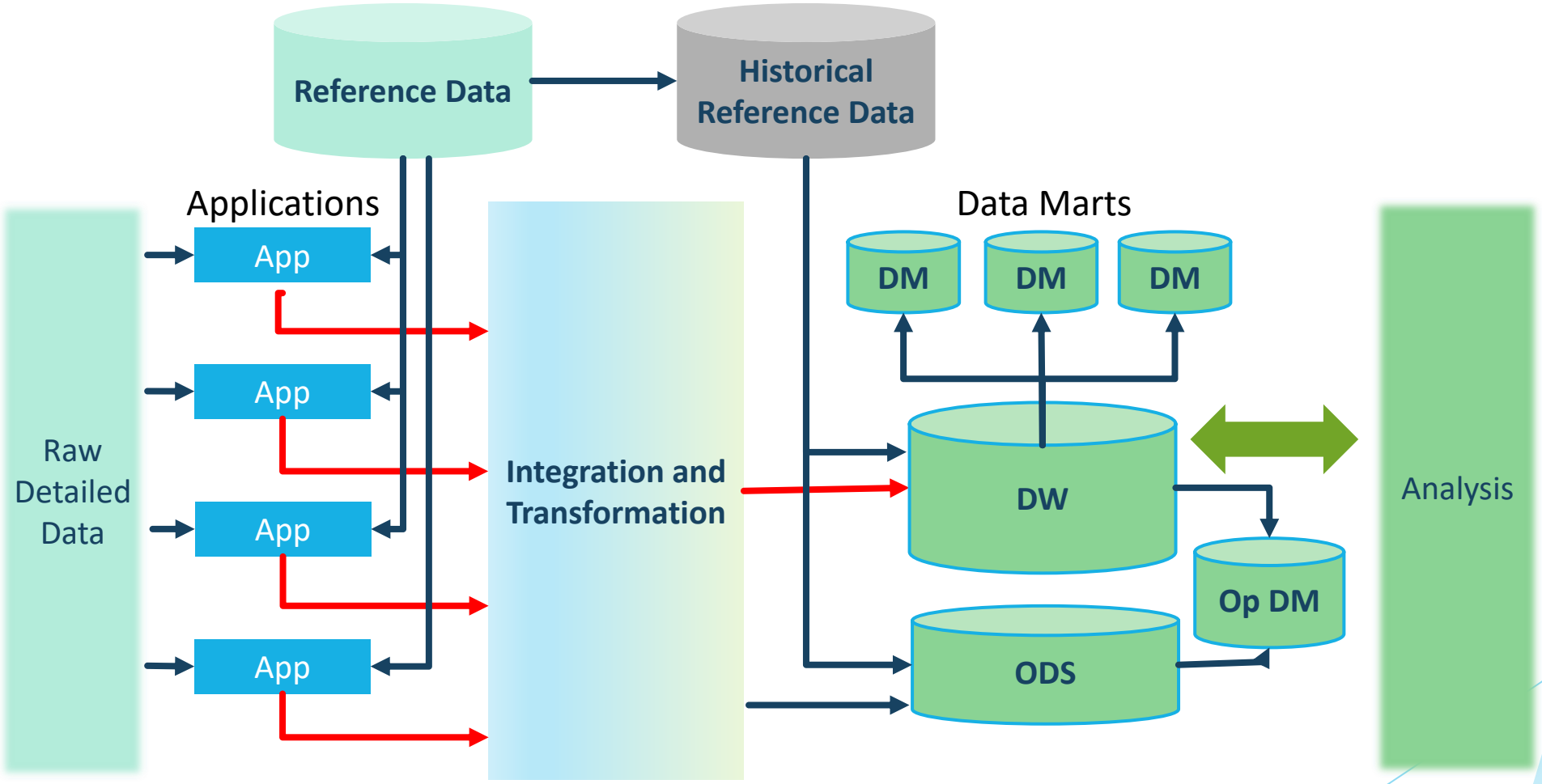
[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

5. Corporate Information Factory (Inmon) – cont'd

- Relational Database Model
- Data Warehouse / Data Marts
    - Staging Area
    - Reference Data, Master Data
    - **Integration and Transformation**
    - Operational Data Store (OD)
    - Operational Data Mart (OpDM)
    - Operational Reports

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

The Corporate Information Factory



[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

6. Dimensional Data Warehouse (Kimball)

- The DW encompasses all components in the data staging and data presentation areas
- **Chess Pieces view of DW/BI architecture**:

    1. Operational source system
    2. Data staging area
    3. Data presentation area
    4. Data access tools

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

- Essential concepts – cont'd

1. **Conformed Dimensions**

   - Dimension tables **conform** when attributes in separate dimension tables have the same column names and domain contents[2]

   - Conformed dimensions allow facts to be categorized in the same way across multiple fact tables, and data marts, ensuring consistent reporting, analytics across the enterprise[3]

Date is a common conformed dimension because its attributes (day, week, month, quarter, year, etc.) have the same meaning when joined to any fact table.

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

[2] https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/conformed-dimension/

[3] https://searchdatamanagement.techtarget.com/

- Essential concepts – cont'd

  2. **Star Schema**

     - Dimensional models are comprised of

       1. **Facts or Measures**, which contain **quantitative** data about the business process

       2. **Dimensions**, which store descriptive attributes (nouns) related to fact data

     - Star Schema comes from a model in which one Fact table joins with many Dimension tables, and when view as a diagram, appears as a Star

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

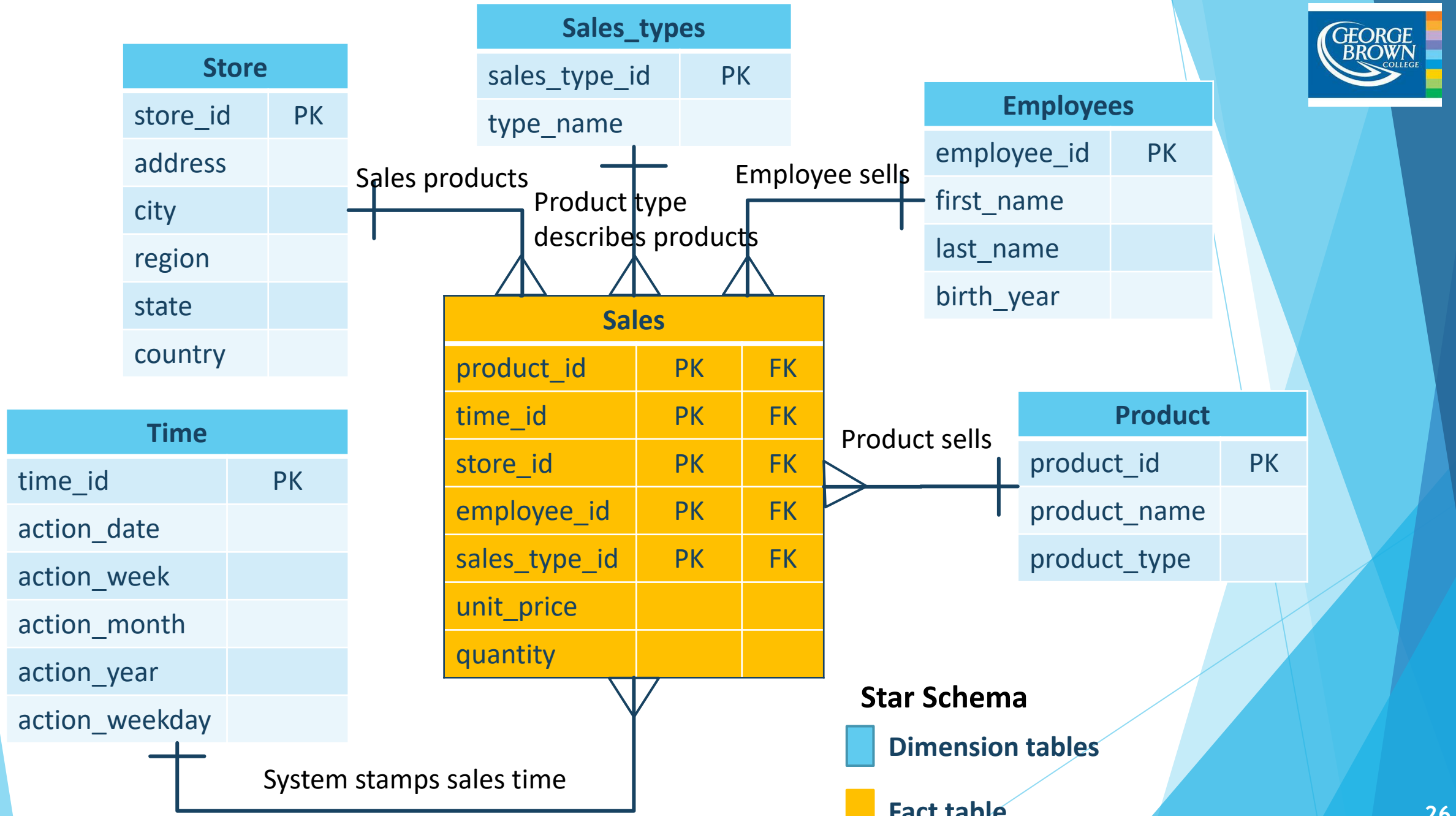
- Essential concepts – cont'd

2. Star Schema [1]

- **Benefits of Star Schemas**
  - Queries are simpler, Easier business insights reporting
  - Better-performing queries
  - Systems can use star schema to build OLAP cubes

- **Challenges of Star Schemas**
  - Decreased data integrity
  - Less capable of handling complex queries

[1] https://www.xplenty.com/blog/snowflake-schemas-vs-star-schemas-what-are-they-and-how-are-they-different/


Store
store_id PK
address
city
region
state
country
Sales_types
sales_type_id PK
type_name
Employees
employee_id PK
first_name
last_name
birth_year
Sales products
Product type describes products
Employee sells
Sales
product_id PK FK
time_id PK FK
store_id PK FK
employee_id PK FK
sales_type_id PK FK
unit_price
quantity
Product
product_id PK
product_name
product_type
Product sells
Time
time_id PK
action_date
action_week
action_month
action_year
action_weekday
System stamps sales time
Star Schema
Dimension tables
Fact table

- Essential concepts – cont'd

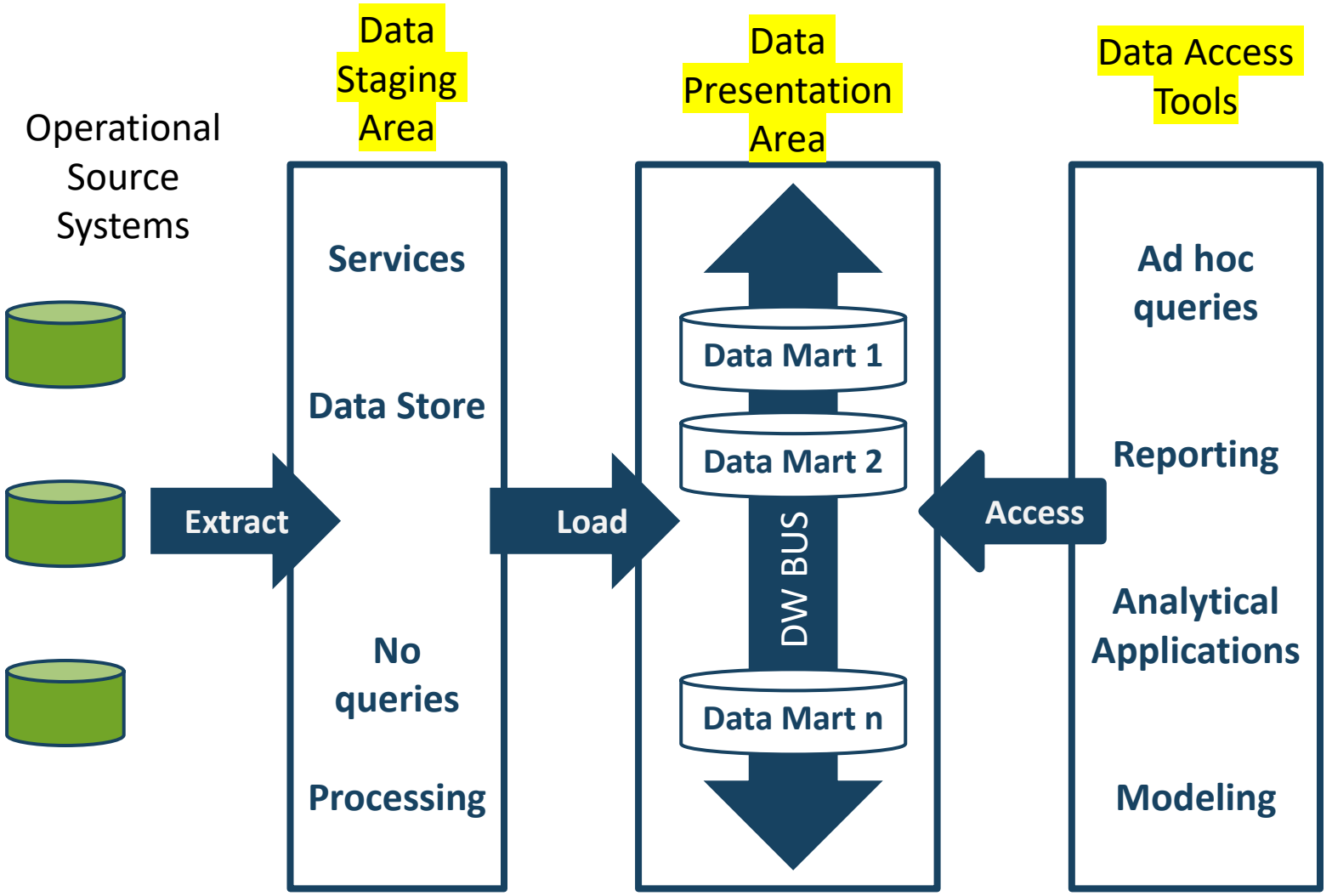
  2. **Snowflake Schema** [1]

     - **Benefits** of Snowflake Schemas
       - Compatible with many OLAP database modeling tools
       - Saves on data storage requirements

     - **Challenges** of Snowflake Schemas
       - Complex data schemas
       - Slower at processing cube data
       - Lower data integrity levels

[1] https://www.xplenty.com/blog/snowflake-schemas-vs-star-schemas-what-are-they-and-how-are-they-different/

Kimball's DW Chess Pieces view of DW/BI



[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

- Essential concepts – cont'd

3. DW Bus

- Multiple Fact tables will share the common, or conformed dimensions via a "DW Bus"

Dimensions

Processes	Subject Areas				
	Date	Product	Store	Vendor	Warehouse facility
Sales	X	X	X		
Inventory	X	X	X	X	X
Orders	X	X		X	
Conformed Dimension Candidate	YES	YES	YES	YES	NO

Fact tables

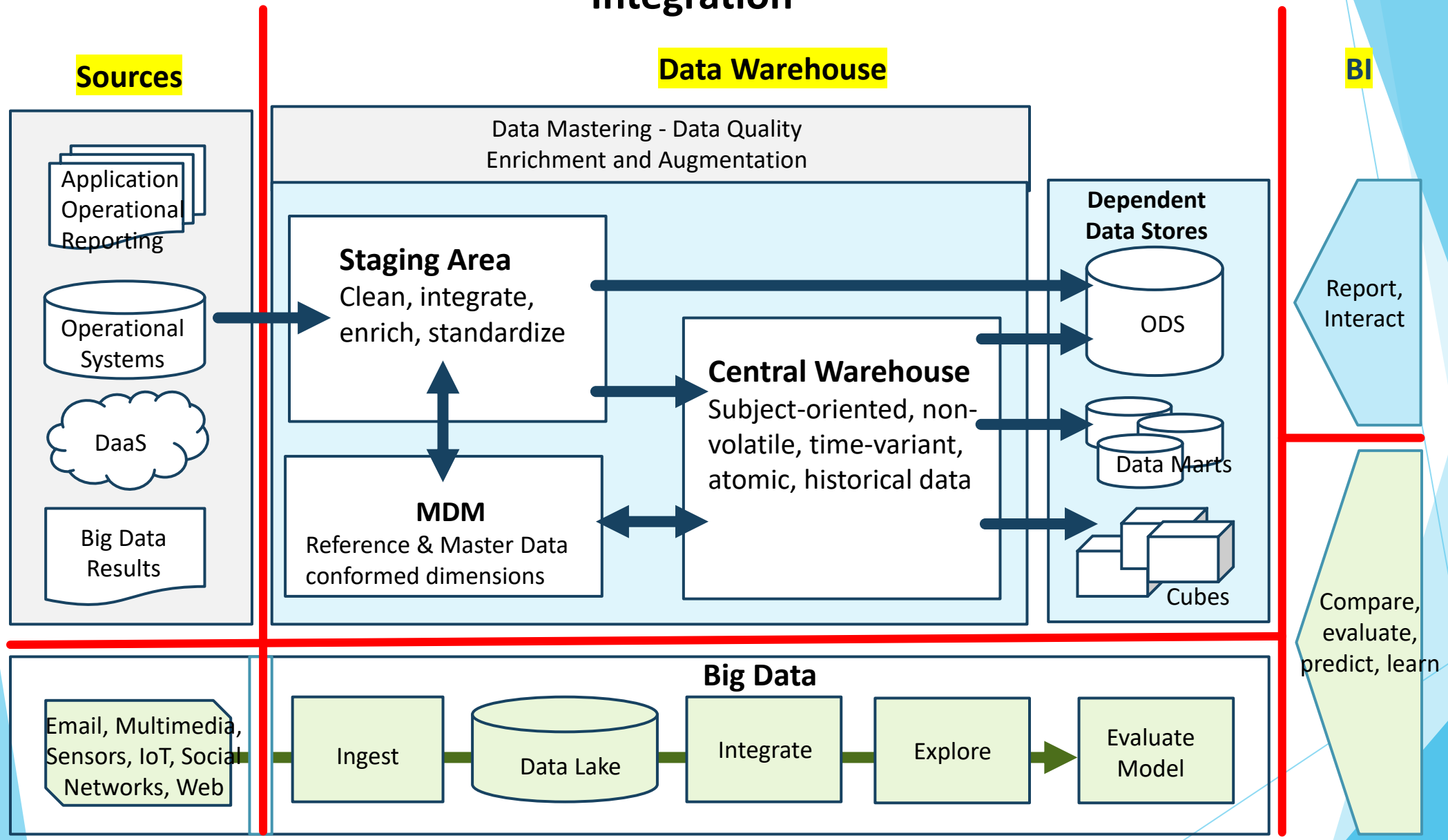
[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

7. DW Architecture Component

- Source Systems
- Data Integration

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Integration



[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

7. DW Architecture Component – cont'd

- Data Storage Areas
  - Staging Area
  - Reference and Master Data conformed dimensions
  - Central Warehouse
- Operational Data Storage (ODS)
- Data Marts
- Cubes

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

8. Type of Load Processing

- Historical data
- Ongoing updates
  - Batch Change Data Capture
  - Near-real-time and Real-time

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Activities

a. Understand requirements
b. Define and maintain the DW and BI architecture
c. Develop the DW and Data Marts
d. Populate the DW
e. Implement the BI portfolio
f. Maintain data products

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

2. ACTIVITIES

a) Understand Requirements

b) Define and Maintain the DW/BI Architecture

- Define DW/BI Technical Architecture
- Define DW/BI Management Processes

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

2. ACTIVITIES – cont'd

c) Develop the Data Warehouse and Data Marts

- Map Sources to Targets
- Remediate and Transform data

d) Populate the Data Warehouse

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

2. ACTIVITIES – cont'd

e) Implement the Business Intelligence Portfolio

- Group Users According to Needs
- Match Tools to User Requirements

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

2. ACTIVITIES – cont'd

f) Maintain Data Products

- Release Management
- Manage Data Product Development Lifecycle
- Monitor and Tune Load Processes
- Monitor and Tune BI Activity and Performance

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

3. TOOLS AND TECHNIQUES

a) Metadata Repository
   - Data Dictionary / Glossary
   - Data and Data Model Lineage

b) Data Integration Tools

c) Business Intelligence Tools Types
   - Operational Reporting
   - Business Performance Management

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

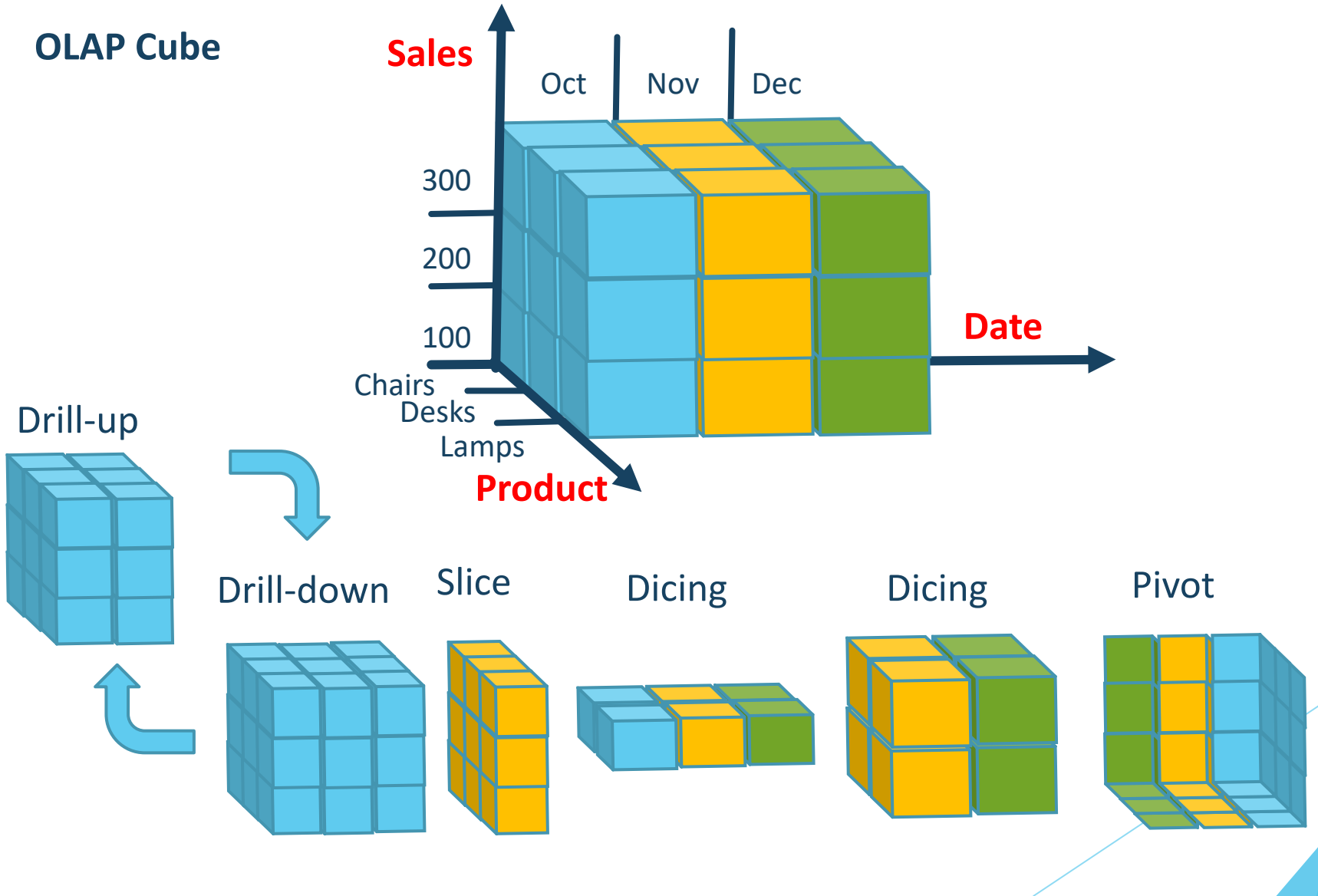
3. TOOLS AND TECHNIQUES

c) Business Intelligence Tools Types – cont'd

- Operational Analytic Applications
  - Multi-dimensional Analysis – OLAP
    - Slice
    - Dice
    - Drill down/up
    - Pivot

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

- Multi-dimensional Analysis – OLAP



4. TECHNIQUES

- Prototypes to Drive Requirements
- Self-service BI
- Audit Data that can be Queried

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

5. IMPLEMENTATION GUIDELINES

- Readiness Assessment / Risk Assessment
- Release Roadmap
- Configuration Management
- Organizational and Cultural Change

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

6. DW/BI GOVERNANCE

- Enabling Business Acceptance
  - Conceptual Model
  - Data Quality feedback loop
  - End-to-end Metadata
  - End-to-end verifiable data lineage
- Customer / User Satisfaction
- SLA
- Reporting Strategy

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

6. DW/BI GOVERNANCE – Cont'd

- Metrics
  - Usage Metrics
  - Subject Area Coverage Percentages
  - Response and Performance Metrics (response time)

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA